Package ‘cordillera’

January 17, 2023

Title Calculation of the OPTICS Cordillera

Version 1.0-0

Date 2023-01-16

Author Thomas Rusch [aut, cre] (https://orcid.org/0000-0002-7773-2096),
Patrick Mair [ctb],
Kurt Hornik [ctb] (https://orcid.org/0000-0003-4198-9911)

Maintainer Thomas Rusch <thomas.rusch@wu.ac.at>

Description Functions for calculating the OPTICS Cordillera. The OPTICS Cordillera measures
the amount of 'clusteredness' in a numeric data matrix within a distance-density based framework for a given minimum number of points comprising a cluster, as described in Rusch, Hornik, Mair (2018) <doi:10.1080/10618600.2017.1349664>. There is an R native version with methods for printing, summarizing, and plotting the result.

Depends R (>= 3.1.2),

SystemRequirements

Imports dbscan

Suggests cluster, scatterplot3d, MASS, R.rsp

VignetteBuilder R.rsp

License GPL-2 | GPL-3

LazyData true

URL https://r-forge.r-project.org/projects/stops/

RoxygenNote 7.2.3

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2023-01-17 10:50:03 UTC
**R topics documented:**

- cordillera-package
- CAClimateIndicatorsCountyMedian
- cordillera
- oldcordilleraplot
- plot.cordillera
- print.cordillera

---

**Description**

A package for calculating the OPTICS Cordillera. The package contains various functions, methods and classes for calculating and plotting the OPTICS Cordillera and an interface to ELKI’s OPTICS.

**Details**

The stops package provides these main functions:

- cordillera() ... OPTICS Cordillera using dbscan OPTICS implementation

Methods: For most of the objects returned by the high-level functions S3 classes and methods for standard generics were implemented, including print, summary, plot.

**References:**


**Authors:** Thomas Rusch

**Maintainer:** Thomas Rusch

**Examples**

data(CAClimateIndicatorsCountyMedian)

res<-princomp(CAClimateIndicatorsCountyMedian[,3:52])
res
summary(res)

library(scatterplot3d)
scatterplot3d(res$scores[,1:3])

irisrep3d<-res$scores[,1:3]
irisrep2d<-res$scores[,1:2]
#OPTICS in dbscan version
library(dbscan)
ores<-optics(irisrep2d,minPts=15,eps=100)
plot(ores)

#OPTICS cordillera for the 2D representation
cres2d<-cordillera(irisrep2d,minpts=15)
cres2d
summary(cres2d)
plot(cres2d)

cres3d<-cordillera(irisrep3d,minpts=15)
cres3d
summary(cres3d)
plot(cres3d)

---

**Climate Change Indicators of Californian Counties**

**Description**

A dataset containing observed and projected indicators of climate change related natural hazards for 58 Californian counties. The values are actually the medians of the predicted distribution over spatial measurement points. It is a compiled data set from three sources and that has been aggregated to the county level. The projected data were derived under two different IPCC climate change scenarios (A2, the high emission scenario and B1, the moderate emission scenario). It further contains the county value of the California social vulnerability index.

**Format**

A data frame with 58 rows and 52 variables

- **county** The county name identifier.
- **vuln_CA** The vulnerability index of Cooley et al. (2012).
- **degFB1** County average 95th percentile daily maximum temperature in Fahrenheit from May 1 to September 30 over the historical period (1971-2000) under the climate scenario B1. These are averaged values for 4 different climate models. The source was Table 7 of Cooley et al. (2012).
- **heatB1_71_00** Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 1971-2000. Projections are based on the B1 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).
- **heatB1_10_39** Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2010-2039. Projections are based on the B1 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).
heatB1_40_69 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2040-2069. Projections are based on the B1 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

heatB1_70_99 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2070-2099. Projections are based on the B1 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

degFA2 County average 95th percentile daily maximum temperature in Fahrenheit from May 1 to September 30 over the historical period (1971-2000) under the climate scenario A2. These are averaged values for 4 different climate models. The source was Table 7 of Cooley et al. (2012).

heatA2_71_00 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 1971-2000. Projections are based on the A2 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

heatA2_10_39 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2010-2039. Projections are based on the A2 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

heatA2_40_69 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2040-2069. Projections are based on the A2 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

heatA2_70_99 Projected average number of days where the daily maximum temperature exceeds the high-heat threshold (see above) over period 2070-2099. Projections are based on the A2 scenario and are averaged for four downscaled climate models. The source was Table 7 of Cooley et al. (2012).

flood_2000 The percentage of a county’s census block area vulnerable to unimpeded coastal flooding under baseline conditions (2000). The raw data were obtained from Heberger et al. (2009). From the census block areas we computed an area-weighted percentage for each county.

flood_2100 The projected percentage of a county’s census block area vulnerable to unimpeded coastal flooding with a 1.4-meter (55-inch) sea-level rise (projected for 2100). The raw data were obtained from Heberger et al (2009). From the census block areas we computed an area-weighted percentage for each county.

basfA2_2000 The median aggregated CCSM3 observed or projected annual baseflow for year 2000 under scenario A2 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfA2_2039 The median aggregated CCSM3 observed or projected annual baseflow for year 2039 under scenario A2 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfA2_2069 The median aggregated CCSM3 observed or projected annual baseflow for year 2069 under scenario A2 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).
basfA2_2099 The median aggregated CCSM3 observed or projected annual baseflow for year 2099 under scenario A2 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfB1_2000 The median aggregated CCSM3 observed or projected annual baseflow for year 2000 under scenario B1 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfB1_2039 The median aggregated CCSM3 observed or projected annual baseflow for year 2039 under scenario B1 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfB1_2069 The median aggregated CCSM3 observed or projected annual baseflow for year 2069 under scenario B1 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

basfB1_2099 The median aggregated CCSM3 observed or projected annual baseflow for year 2099 under scenario B1 by county (past years are observed, future years are projected). The source of the raw data was California Energy Commission (2008).

evapA2_2000 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2000 under scenarios A2 by county. The source of the raw data was California Energy Commission (2008).

evapA2_2039 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2039 under scenarios A2 by county. The source of the raw data was California Energy Commission (2008).

evapA2_2069 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2069 under scenarios A2 by county. The source of the raw data was California Energy Commission (2008).

evapA2_2099 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2099 under scenarios A2 by county. The source of the raw data was California Energy Commission (2008).

evapB1_2000 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2000 under scenarios B1 by county. The source of the raw data was California Energy Commission (2008).

evapB1_2039 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2039 under scenarios B1 by county. The source of the raw data was California Energy Commission (2008).

evapB1_2069 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2069 under scenarios B1 by county. The source of the raw data was California Energy Commission (2008).

evapB1_2099 The median aggregated Community Climate System Model v.3 (CCSM3) projected annual actual evapotranspiration for year 2099 under scenarios B1 by county. The source of the raw data was California Energy Commission (2008).


prcpA2_2039 The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

prcpA2_2069 The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

prcpB1_2000  The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

prcpB1_2039  The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

prcpB1_2069  The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

prcpB1_2099  The median aggregated CCSM3 projected annual precipitation for year 2000 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

smclA2_2000  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2000 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

smclA2_2039  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2039 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

smclA2_2069  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2069 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

smclA2_2099  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2099 under scenario A2 by county. The source of the raw data was California Energy Commission (2008).

smclB1_2000  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2000 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

smclB1_2039  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2039 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

smclB1_2069  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2069 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

smclB1_2099  The median aggregated CCSM3 projected annual fractional moisture in the entire soil column for year 2099 under scenario B1 by county. The source of the raw data was California Energy Commission (2008).

fireA2_2020  The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in year 2020 under scenarios A2. The source of the raw data was California Energy Commission (2008).

fireA2_2050  The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in year 2050 under scenarios A2. The source of the raw data was California Energy Commission (2008).

fireA2_2085  The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in
year 2085 under scenarios A2. The source of the raw data was California Energy Commission (2008).

**fireB1_2020** The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in year 2020 under scenarios B1. The source of the raw data was California Energy Commission (2008).

**fireB1_2050** The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in year 2050 under scenarios B1. The source of the raw data was California Energy Commission (2008).

**fireB1_2085** The median aggregated Centre National de Recherches Meteorologiques (CNRM) projected annual wildfire risk (observing 1 or more fires in the next 30 years) for each county in year 2085 under scenarios B1. The source of the raw data was California Energy Commission (2008).

Details

Overall there are 50 indicators of natural hazard, one indicator of social vulnerability and 1 identifier of the county which were:

Source


---

**cordillera**

*The OPTICS Cordillera*

Description

Calculates the OPTICS Cordillera as described in Rusch et al. (2017). Based on optics in dbscan package.

Usage

```r
cordillera(
  X,
  q = 2,
  minpts = 2,
  epsilon,
  distmeth = "euclidean",
```


dmax = NULL,
rang,
digits = 10,
scale = FALSE,
...
)

Arguments

X numeric matrix or data frame representing coordinates of points, or a symmetric matrix of distance of points or an object of class dist. Passed to optics, see also there.
q The norm used for the Cordillera. Defaults to 2.
minpts The minimum number of points that must make up a cluster in OPTICS (corresponds to k in the paper). It is passed to optics where it is called minPts. Defaults to 2.
epsilon The epsilon parameter for OPTICS (called epsilon_max in the paper). Defaults to 2 times the maximum distance between any two points.
distmeth The distance to be computed if X is not a symmetric matrix (those from dist are available) or a dist object (otherwise ignored). Defaults to Euclidean distance.
dmax The winsorization value for the highest allowed reachability. If used for comparisons this should be supplied. If no value is supplied, it is NULL (default), then dmax is taken from the data as minimum of epsilon or the largest reachability.
rang A range of values for making up dmax. If supplied it overrules the dmax parameter and rang[2]-rang[1] is returned as dmax in the object. If no value is supplied rang is taken to be (0, dmax) taken from the data. Only use this when you know what you’re doing, which would mean you’re me (and even then we should be cautious).
digits The precision to round the raw Cordillera and the norm factor. Defaults to 10.
scale Should X be scaled if it is an asymmetric matrix or data frame? Can take values TRUE or FALSE or a numeric value. If TRUE or 1, standardisation is to mean=0 and sd=1. If 2, no centering is applied and scaling of each column is done with the root mean square of each column. If 3, no centering is applied and scaling of all columns is done as X/max(standard deviation(allcolumns)). If 4, no centering is applied and scaling of all columns is done as X/max(rmsq(allcolumns)). If FALSE, 0 or any other numeric value, no standardisation is applied. Defaults to FALSE.
...

Value

A list with the elements

- $raw... The raw cordillera
- $norm... The normalization constant
- $normfac... The normalization factor (the number of times that dmax is taken)
• $dmaxe... The effective maximum distance used for maximum structure (either dmax or epsilon or rang[2]-rang[1]).
• $normed... The normed cordillera (raw/norm)
• $optics... The optics object

Warning
It may happen that the (normed) cordillera cannot be calculated properly (e.g. division by zero, infinite raw cordillera, q value to high etc.). A warning will be printed and the normed Cordillera is either 0, 1 (if infinity is involved) or NA. In that case one needs to check one or more of the following: reachability values returned from optics, minpts, eps, the raw cordillera, dmax and the normalization factor normfac.

Examples

```r
data(iris)
res<-princomp(iris[,1:4])
#2 dim goodness-of-clusteredness with clusters of at least 2 points
#With a matrix of points
cres2<-cordillera(res$scores[,1:2])
cres2
summary(cres2)
plot(cres2)

#with a dist object
dl0 <- dist(res$scores[,1:2],"maximum") #maximum distance
cres0<-cordillera(dl0)
cres0
summary(cres0)
plot(cres0)

#with any symmetric distance/dissimilarity matrix
dl1 <- cluster::daisy(res$scores[,1:2],"manhattan")
cres1<-cordillera(dl1)
cres1
summary(cres1)
plot(cres1)

#4 dim goodness-of-clusteredness with clusters of at least 20 points for PCA
cres4<-cordillera(res$scores[,1:4],minpts=20,epsilon=13,scale=3)
#4 dim goodness-of-clusteredness with clusters of at least 20 points for original data

cres<-cordillera(iris[,1:4],minpts=20,epsilon=13,dmax=cres4$dmaxe,scale=3)
#There is more clusteredness for the original result
summary(cres4)
scale(cres)
plot(cres4) #cluster structure only a bit intelligible
plot(cres) #clearly two well separated clusters
```

###############################################################################
# Example from Rusch et al. (2018) with original data, PCA and Sammon mapping#

#data preparation
data(CAClimateIndicatorsCountyMedian)
sovisel <- CAClimateIndicatorsCountyMedian[,c(1,2,4,9)]
#normalize to [0,1]
sovisel <- apply(sovisel,2,function(x) (x-min(x))/(max(x)-min(x)))
rownames(sovisel) <- CAClimateIndicatorsCountyMedian[,1]
dis <- dist(sovisel)

#hyper parameters
dmax=1.22
q=2
minpts=3

#original data directly
cdat <- cordillera(sovisel,distmeth="euclidean",minpts=minpts,epsilon=10,q=q,
scale=0)
#equivalently
#dis2=dist(sovisel)
#cdat2 <- cordillera(dis2,minpts=minpts,epsilon=10,q=q,scale=FALSE)

#PCA in 2-dim
pca1 <- princomp(sovisel)
pcas <- scale(pca1$scores[,1:2])
cpca <- cordillera(pcas,minpts=minpts,epsilon=10,q=q,dmax=dmax,scale=FALSE)

#Sammon mapping in 2-dim
sam <- MASS::sammon(dis)
samp <- scale(sam$points)
csam <- cordillera(samp,epsilon=10,minpts=minpts,q=q,dmax=dmax,scale=FALSE)

#results
cdat
cpca
csam

par(mfrow=c(3,1))
plot(cdat)
plot(cpca)
plot(csam)
par(mfrow=c(1,1))
Description

Plots the reachability plot and adds the cordillera to it (as a line). In this plot the cordillera is proportional to the real value.

Usage

oldcordilleraplot(
  x,
  colbp = "lightgrey",
  coll = "black",
  liwd = 1.5,
  legend = FALSE,
  ylim,
  ...
)

Arguments

x an object of class cordillera
colbp color of the barplot.
coll color of the cordillera line
liwd width of the cordillera line
legend draw legend
ylim ylim for the barplots
... additional arguments passed to barplot or lines

plot.cordillera

Plot method for OPTICS Cordilleras

Description

Plots the reachability plot and adds the cordillera to it (as a line). In this plot the cordillera is proportional to the real value.

Usage

## S3 method for class 'cordillera'
plot(  
  x,
  colbp = "lightgrey",
  coll = "black",
  liwd = 1.5,
  legend = FALSE,
  ylim,
  ...
)

print.cordillera

Arguments

- `x` an object of class "cordillera"
- `colbp` color of the barplot.
- `coll` color of the cordillera line
- `liwd` width of the cordillera line
- `legend` draw legend
- `ylim` ylim for the barplots
- `...` additional arguments passed to barplot or lines

Description

Prints the raw and normalized OPTICS Cordillera

Usage

```r
## S3 method for class 'cordillera'
print(x, ...)
```

Arguments

- `x` an object of class optics
- `...` additional arguments passed to print
Index

- clustering
  cordillera, 7
- multivariate
  cordillera, 7

CAClimateIndicatorsCountyMedian, 3
cordillera, 7
cordillera-package, 2
dist, 8
oldcordilleraplot, 10
optics, 8
plot.cordillera, 11
print.cordillera, 12